

Developing data literacy: How data services and data fellowships are creating data skilled social researchers

Vanessa Higgins and Jackie Carter¹

Abstract

This paper describes two successful approaches to data literacy training within the UK and the synergies and collaborations between these two programmes. The first is a data literacy training programme, being delivered by the UK Data Service, which focuses on training in basic data literacy skills. The second is a Data Fellows programme that has been developed to help undergraduate social science students gain real-world experience by applying their classroom skills in the workplace. The paper also discusses next steps in the global development of data literacy skills via the EmpoderaData project, which is trialling the Data Fellows programme in Latin America.

Keywords

data literacy, information literacy, data services, data skills, sustainable development goals

Introduction

A lack of quantitative data skills among social scientists in the UK has been recognised for over twenty years by government, businesses and research funders (Carter et al, 2021c; MacInnes, 2009; UK DCMS Policy, 2021). The importance of this has become more apparent during the coronavirus pandemic when data literacy skills have been critical for research capabilities. As the world becomes increasingly data-rich, we need more data-literate social scientists to contribute to our social and economic understanding locally, nationally and globally. This is particularly important for countries to deliver against the Sustainable Development Goal (SDGs), which rely heavily on quantitative data (United Nations, 2015; Higgins et al, 2019).

There is a wealth of quantitative socio-economic data available for secondary reuse in national data archives/services. For instance, there are national and cross-national government survey data, Census data and longitudinal data. These data are incredibly rich research resources, a lot of which is collected by Government departments, and could be used much more widely for social research and policymaking if quantitative data literacy skills were more widespread among social researchers. This paper describes two successful approaches to data literacy training within the UK, the synergies and collaborations between these two programmes, and how one is being trialled in Latin America. The first is a data literacy training programme being delivered by the [UK Data Service](#), the second is a Data Fellows programme that has been developed to help undergraduate social science students gain real-world experience by applying their classroom skills in the workplace.

Throughout this paper we use the term 'data literacy' to describe the broad set of skills that are required to enable data that are hosted in data services and archives (such as the UK Data Service) to be discovered, critically evaluated and deployed in social research. We follow Data-Pop Alliance's approach in taking data literacy to mean "the desire and ability to constructively engage in society through or about data" (Data-Pop Alliance, 2015). As such, data literacy is a broad and pragmatic term which intersects with and builds on other literacies, such as information and statistical literacies. Data literacy as used here therefore requires conceptual and mechanical understanding in order to be able to *do* data analysis. Moreover, we focus here on the use of quantitative data and analysis, and the Data Fellows programme is specifically aimed at increasing skills in this area to enable students to critically evaluate and use numerical data (usually but not always statistical data) through data-driven research projects. For further information about how we apply this to the Sustainable Development Goals context see Carter et al (2021c).

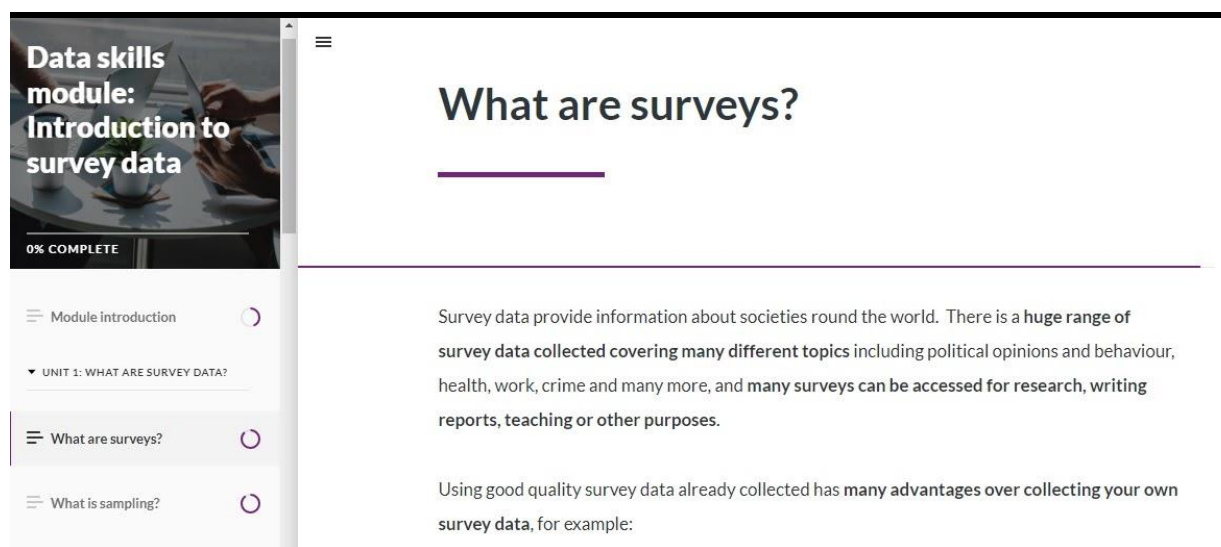
Data literacy training via the UK Data Service

Many data archives or data services have data literacy training programmes to encourage and enable the use of the data that they make available for reuse, for example, the [UK Data Service](#) (UKDS), [GESIS Leibniz Institute for the Social Sciences](#), the [Inter-University Consortium for Political and Social Research](#) and [Consortium of European Social Science Data Archives](#). Such programmes are vital to ensure the use of the data held by these archives/services because they provide training to use the specific data available to researchers, rather than generic statistical methodology courses.

The UKDS data literacy training programme provides a combination of training events and web-based on-demand training materials targeted at an introductory level of data literacy training. The programme includes online workshops with short practical sessions on topics such as *Getting Started with Secondary Analysis* and *Data in the Spotlight* workshops to introduce different data types, as well as other formats for training such as *coding demos*, where code is worked through line-by-line and *drop-in sessions* where attendees can drop in to meet experts from the UKDS to get general advice on using data. On-demand training materials are also available from the UKDS website and include short instructional videos and written guidance on topics such as Weighting Survey Data and Getting Started with SPSS. The programme also includes a series of interactive Data Skills Modules (Figure 1) designed for anyone who wants to start using secondary data, with no prior knowledge required. Learners can follow the modules in their own time, dipping in and out when needed, and they receive a certificate of completion at the end of each module.

The UK Data Service training events and on-demand training materials are targeted at a basic level which encourages use of the data among those who have no/little formal data literacy training. This model encourages use among those who may want to dip their toes in the water but do not want to study on a more formal course. Another attractive element for newcomers is that the training is free of charge and mostly online so there is no financial cost involved. This also encourages wider participation from those who would not normally be able to access such training, such as voluntary sector organisations, local government and students.

Figure 1: [Data Skills Module](#) example



The screenshot shows a web-based learning interface. On the left is a sidebar with a dark header that reads 'Data skills module: Introduction to survey data' and '0% COMPLETE'. Below this, the sidebar lists 'Module introduction' and 'UNIT 1: WHAT ARE SURVEY DATA?' with a dropdown arrow. Under the unit, two items are listed: 'What are surveys?' and 'What is sampling?', each with a circular progress indicator. The main content area has a white background and a purple horizontal line under the title 'What are surveys?'. The text below the title reads: 'Survey data provide information about societies round the world. There is a huge range of survey data collected covering many different topics including political opinions and behaviour, health, work, crime and many more, and many surveys can be accessed for research, writing reports, teaching or other purposes.' Below this, it says: 'Using good quality survey data already collected has many advantages over collecting your own survey data, for example:'.

The basic, introductory nature of the UKDS training events and materials makes them very popular with over 7000 attendees at events per year and in excess of 7000 views, per month, of the materials

on the UKDS YouTube channel. Examples of the benefits of the training events are highlighted in the feedback we received from attendees:

'It has given me a better idea of what data is already there and stimulated embryonic research ideas! Thank you.' (Research Fellow)

'The presentation was extremely useful. It covered a lot of relevant content and contained lots of good links for further reading. The Q & A session was helpful to understand how researchers could apply these rules in practice.' (Research Technology Specialist)

'Bravo! I am a trained computer scientist (hugely comfortable with assumptions, definition, classification, hierarchy) working in computational social science and striving hard to embrace and learn abstraction skills. Have never seen anyone summarise this so well before! I feel seen and heard.' (Computer Scientist)

'This will mean I can make thematic maps myself, without having to outsource this to colleagues or contractors.' (Anon)

'Really helpful introduction and prompted me to sign up to the next one, which feels like a natural follow-on. Thank you!' (Research Associate)

Likewise, the utility of the Data Skills Modules is highlighted in comments we received from learners on completion of the modules.

'Extremely useful, very thorough and engaging examples. Layout is very easy to use and aesthetically pleasing.' (Undergraduate student)

'I really enjoyed it and I have encouraged my students to complete the module as part of the learning content. A good, all-round introduction.' (Lecturer at Higher Education Institution)

'This module was very useful. Further modules would be interesting on how to use different tests for survey data (ANOVA, t-tests, etc).' (Not-for-profit sector user)

Data skills training via the University of Manchester Data Fellows programme

In 2013, The University of Manchester, UK, established the Q-Step Centre. 'Q-Step' was a government funded initiative to support the development of quantitatively skilled social science students across UK universities, to help '(i) create a step change in teaching undergraduate social science students quantitative research skills, and (ii) develop a talent pipeline for future careers in applied social research.' (Carter, 2021a). As part of the focus on application of data skills, the University of Manchester Q-Step (UMQStep) Centre established work placements, which has come to be called the Data Fellowship programme. Since the summer of 2014, three hundred undergraduates have been placed as data fellows into public, private and not-for-profit organisations to carry out data-driven research projects over a two-month period sandwiched between the end of the second and start of the third year of their degree. All students are paid the living wage, ensuring the placements are available to all and not the preserve of those who could afford to do these without payment. As a result, 25% of the placements have been taken up by those from less-advantaged backgrounds or under-represented groups, and 70% have been occupied by females. The programme is therefore addressing not only the need to help social science students acquire data skills practice in the workplace, but is also delivering a diverse talent pipeline into graduate social research careers. The learnings from the programme have been captured in a book on 'Work placements, internships and applied social research' (Carter, 2021b) which helps students find, prepare for, do and reflect on an

internship. Ten case studies and several vignettes are included in the book to illustrate the analytical and research skills, and professional skills, that can be developed through work experience focused on social research, together with examples of early career researchers working in social research organisations.

It is important to note that the students who participate in the data fellowships programme are taught statistics and data analysis skills during the first two years of their undergraduate study. As social science majors they are on degree programmes through which they are studying subjects such as criminology, politics and international relations, sociology or social anthropology. The approach taken at the University has been to embed the teaching of data analysis into these substantive subjects (Buckley et al, 2015). Specific examples of course units teaching data skills can be found in Carter (2021b). Prior to starting placement students are given refresher courses to enable them to enter the workplace reasonably well-prepared for the data-driven project(s) they will work on. The combination of the classroom teaching and the embedding of data into their social science subject means that students will already have encountered real-world data through their education. For example, criminology students will have used nationally representative crime surveys, politics students will have analysed elections data from, for example, the British Elections Study, and sociology students will have worked with data from studies such as the British Social Attitudes survey. Many will also have been exposed to international data sources through courses in their first year, for example the Measuring Inequalities unit - a core first year course - uses the World Values Survey and in the Engaging with Social Research compulsory methods course all students on a BA in Social Science programme will have been introduced to qualitative and quantitative data sources used in their lecturers' research. In most cases they will have been introduced to this data in lectures and then handled the data in PC labs or virtually through practical sessions.

Consequently, Manchester undergraduate social science students will have been exposed to real-world data through their first two years at university, and some will be starting to focus on these skills to specialise on a potential future career in applied social research, whilst others will be keeping their options open and following a broader curriculum. They will have used different software (ranging from Excel to R) and have covered a range of methods (from descriptive statistics to bivariate analysis) by the time they take up a data fellowship.

Creating synergies between the two programmes

The two programmes outlined above are both led by the University of Manchester. With a proud history of supporting access to quantitative data for empirical research, and strengths in teaching data skills, the university was well-placed to create successful synergies between the two programmes. Moreover, with its third strategic goal of 'Social Responsibility' and being a world leader in the impact of its research and teaching as measured against the UN's Sustainable Development Goals (Times Higher, 2021), the University of Manchester is a natural home of the development of data literacy and data skills in the social sciences.

The two case studies below evidence how Manchester undergraduates have, through the data fellowship programme, worked with the UK Data Service to enhance their professional and analytical skills and have used this learning in the subsequent steps into their careers. Many others have also benefited from the connection between the two programmes, and the two included here are merely selected to illustrate the types of data-driven projects that enable students to springboard into careers that value the data literacy they acquire, coupled with the social science degree subject that they study.

In the first case study the undergraduate was placed as a data fellow with a not-for-profit organisation. The experience opened her eyes to a career in supporting data uses and she now works with the UK

Data Service. In the second example the student undertook her data fellowship with the UK Data Service, helping to update the Index of Multiple Deprivation with the latest census data. She now works as a government analyst.

Case study 1: Alle Bloom

As a Data Fellow Alle completed a placement with Respect helpline (a domestic abuse helpline service), then upon graduating completed a Master's in Social Research Methods and Statistics at the University of Manchester. She is now a Research Associate for the UK Data Service where she is teaching data literacy skills and helping others to extract value from data.

Alle says:

'My data fellowship took me from looking at numbers on my university computer screen, to understanding the people and mechanisms behind them in the 'real-world'. I've always learned better by 'doing' and after experiencing the whole process of taking data from collection through to reports that had a tangible impact, I realised this is what I wanted to do. I now work for the UK Data Service, helping others with their research and teaching, and still continue to learn more about our data and the research process everyday.'

Case study 2: Klara Valentova

Klara was highly motivated by her first- and second-year course units to learn more about how data indices are created. She secured a data fellowship with the UK Data Service, helping them to update the Carstairs Index of Deprivation that had been developed in the 1980s. As part of her fellowship she wrote a blog post (<https://blog.ukdataservice.ac.uk/meet-our-interns-klara/>) from which this extract is adapted:

'Carstairs ... comprises four indicators from the Census, which relate to material deprivation (overcrowding, male unemployment, low social class and lack of car ownership). Some of these variables are a bit outdated, and so we include other indicators, which are more up to date. ... [including] total unemployment (female and male combined) in our calculations as there are [many] more women in the labour force than there were nearly 40 years ago.'

She went on to use the findings (and outputs) from her placement in her final year dissertation, took that into her graduate study and is now working at the Office for National Statistics. The data fellowship opened up a career route for her that she had previously not considered, and she was able to evidence the skills and knowledge she had acquired during her undergraduate study.

Klara's reflections on her time spent as a data fellow are captured here:

'I realised just how powerful data can be. By examining a deprivation index created in the 1980s, it became clear that if outdated data or inappropriate methods are used, the outputs are likely to be biased and in turn lead to inaccurate policy decision making. On the other hand, if data is analysed correctly, it can be a powerful tool to help improve people's lives. I am now working as a government analyst promoting best practice in data collection, analysis and dissemination so we can draw more value from statistics for public good.'

The combination of teaching with real-world data made available through the UK Data Service, and helping students acquire data skills through the workplace, is a powerful one. The success story is in the development of a pipeline of talented, curious, data literate social science graduates who can enter careers in the data professions.

Developing data literacy globally – next steps

The success of the Data Fellows programme has led to interest from education and civic society organisations interested in developing the model in their own countries. Through a collaboration with DataPop Alliance, that resulted from a Data Fellow being placed with an organisation they work closely with (Open Data Watch), we have been able to develop an international dimension to the initiative. We describe the origins and early stages of this research in Carter et. al (2021c) where we discuss the EmpoderaData project which has explored the transferability of the data fellows scheme to Colombia, Mexico and Brazil. The full results from the early stages of the EmpoderaData project are available in Higgins et al (2019).

The EmpoderaData research to date has uncovered a need for basic data skills training (similar to those covered in the UK Data Service data literacy programme) within the three case study countries. Moreover, within Brazil and Colombia, the data fellowship model is perceived as a valuable tool for building basic data literacy capacity to help deliver the SDGs. The notion of hybrid professional teams has emerged which would bring together ‘fellows’ with complementary backgrounds to work collaboratively on SDG-related research projects at host organisations. Examples of hybrid professional teams could be social scientists and data scientists working together or social scientists working alongside STEM graduates (Higgins et al, 2019; Carter et al, 2021c). As a direct result of the EmpoderaData research project, two pilot Data Fellow programmes are being implemented in Colombia and Brazil (Jones et al, 2021; Carter et al 2021c). These early initiatives are a positive step forward and they illustrate that the University of Manchester Data Fellows model can be adapted to different disciplines (business studies and mathematics) and has international relevance and appeal.

Conclusion

There is a wealth of quantitative socio-economic data available for secondary reuse in national data archives/service but there is a lack of quantitative data literacy skills among social researchers, or others who may want to use these data; this is particularly important for the monitoring of the SDGs. Different models of data literacy training such as the UK Data Service data literacy training programme and the University of Manchester Data Fellows programme have created successful synergies to benefit undergraduate social science students’ data literacy training experience. The synergies outlined in this paper are a useful model for other data literacy programmes to apply to strengthen the pipeline of data skills training among undergraduates. The EmpoderaData research project has led to new initiatives to trial Data Fellow Programmes in Brazil and Colombia. If successful, the programme has potential to be expanded to more countries and across different disciplines. This collaboration provides an exciting future research agenda which we will be evaluating carefully.

Acknowledgments

The authors would like to thank Julie Ricard, Valentina Casasbeunas and Emmanuel Letouzé from Data-Pop Alliance for their collaboration on EmpoderaData Phase 1. A huge thanks to the UK Data Service staff members who deliver training events and materials and those who hosted the Data Fellows. Thanks to Klara Valentova and Alle Bloom for their permission to include them as case studies within this paper. Financial support for the research reported here has been thanks to various grants from the UK ESRC (Economic and Research Council), The Nuffield Foundation and the UK Global Challenges Research Fund (GCRF) as well as the University of Manchester who co-fund the data fellowships.

References

- Buckley, J, Brown, M, Thompson, S, Olsen, W & Carter, J 2015, 'Embedding Quantitative Skills into the Social Science Curriculum: Case studies from Manchester', *International Journal of Social Research Methodology*. <https://doi.org/10.1080/13645579.2015.1062624>.
- Carter, J 2021a, 'Developing a future pipeline of applied social researchers through experiential learning: the case of a data fellows programme', *Statistical Journal of the IAOS*, pp. 1-16. <https://doi.org/10.3233/SJI-210844>.
- Carter, J (2021b) *Work Placements, Internships and Applied Social Research*, Sage, London.
- Carter, J, Méndez-Romero, RA, Jones, P, Higgins, V & Samartini, ALS. (2021c). EmpoderaData: Sharing a Successful Work-placement Data Skills Training Model Within Latin America, to Develop Capacity to Deliver the SDG, *Statistical Journal of the IAOS*. <https://doi.org/10.3233/SJI-210842>.
- Data-Pop Alliance. 2015 *Beyond Data Literacy: Reinventing Community Engagement and Empowerment in the Age of Data*. White Paper. http://datapopalliance.org/wp-content/uploads/2015/10/BeyondDataLiteracy_DataPopAlliance_Sept30.pdf.
- Higgins, V, Casasbuenas, V, Ricard, J & Carter, J 2019, *EmpoderaData: Data Literacy Assessment and Sustainable Development Goals Data Gaps: Brazil, Colombia and Mexico*. Data-Pop Alliance. https://datapopalliance.org/wp-content/uploads/2019/11/EMPODERADATAREPORT_final_oct2019.pdf.
- Jones, P, Carter, J, Renken, J & Arbeláez Tobón, M 2021 'Strengthening the Skills Pipeline for Statistical Capacity Development to Meet the Demands of Sustainable Development: Implementing a Data Fellowship Model in Colombia' Digital Development Working Paper Series, 89 edn, University of Manchester, Global Development Institute, Manchester. http://hummedia.manchester.ac.uk/institutes/gdi/publications/workingpapers/di/dd_wp89.pdf.
- MacInnes, J. (2009). *Final report: Proposals to support and improve the teaching of quantitative research methods at undergraduate level in the UK*. Swindon: Economic and Social Research Council.
- Times Higher (2021) *SDGs Global Impact Rankings* https://www.timeshighereducation.com/impactrankings#!/page/0/length/25/sort_by/rank/sort_order/asc/cols/undefined [accessed 19 Jan 2022].
- UK DCMS Policy (2021) *United Kingdom Government Department for Digital, Culture, Media and Sport Policy Paper: National Data Strategy* [Internet] gov.uk 2020 [cited 25 May 2021]. Available from: <https://www.gov.uk/government/publications/uk-national-data-strategy/national-data-strategy>.
- United Nations. *Transforming our world: The 2030 agenda for Sustainable development* [Internet]. Sustainable Development GOALS Knowledge Platform; 2015 [cited 2021 May 20] p. 41. Report No.RES/70/ <https://sustainabledevelopment.un.org/content/documents/21252030%20Agenda%20for%20Sustainable%20Development%20web.pdf>.

Endnotes

¹ Dr Vanessa Higgins, University of Manchester, vanessa.higgins@manchester.ac.uk (corresponding author)

² Professor Jackie Carter, University of Manchester, Jackie.Carter@manchester.ac.uk